

HP-10014772-1

Patent

UNITED STATES PATENT APPLICATION

for

A METHOD FOR PLACEMENT OF DATA FOR VISUALIZATION OF
MULTIDIMENSIONAL DATA SETS USING MULTIPLE PIXEL BAR CHARTS

Inventors:

Ming C. Hao
Umeshwar Dayal
Meichun Hsu
Daniel A. Kiem
Adrian Krug
Julian Ladisch

prepared by:

WAGNER, MURABITO & HAO LLP

Two North Market Street

Third Floor

San Jose, CA 95113

(408) 938-9060

FILED FOR THE U.S. PATENT OFFICE

A METHOD FOR PLACEMENT OF DATA FOR VISUALIZATION OF MULTIDIMENSIONAL DATA SETS USING MULTIPLE PIXEL BAR CHARTS

FIELD OF INVENTION

5 The present invention relates to the field of information visualization. Specifically, the present invention relates to a method for placement of data for visualization of multidimensional data sets using multiple pixel bar charts.

BACKGROUND OF THE INVENTION

10 Pixel-oriented techniques allow for the presentation of large amounts of data having detailed information. To date, many visualization applications utilize of pixel-oriented techniques, such as the spiral technique, the recursive pattern technique and the pixel bar chart technique. The pixel bar chart technique is for presenting the data values directly rather than aggregating the data into a few data values. The
15 approach is to represent each item by a single pixel in the bar chart.

 However, for an analysis of large volumes of data (e.g., large volumes of e-commerce transactions), the visualization of complex data is not sufficient. In order to optimally view and interpret the data it is necessary to optimally place the data. It
20 is not obvious what an optimal placement of pixel should be. The optimal placement of data associated with visualization of massive multi-attribute data presents particular problems.

For one, in large data sets (e.g., dense data sets), limitations exist on pixel visualizations that are capable of showing large amounts of data on a value by value basis without aggregation. Current techniques do not offer optimal arrangement of data for detecting close relationships among the data. Also, current techniques do not offer an optimal arrangement of data for representing multiple attributes.

Current pixel visualization techniques also do not present data such that similar data records are placed close to each other. It is desirable to place similar records close to each other to fully appreciate the information presented by the data. For example, the detection of trends among data requires that similar records be placed in close proximity.

Furthermore, current pixel visualization techniques do not arrange data in a logical sequence. Currently techniques provide insufficient placement to resolve the locality and ordering constraints. As such, data is not categorized, and is more difficult to understand.

Also, limitations exist such that current pixel visualization techniques are restricted with respect to the number of the data attributes that may be used. Current techniques are typically limited to either four or six attributes. Since it is not uncommon to have data sets comprised of greater than 10 attributes, this limitation

can severely restrict optimal viewing and interpretation of the data.

Accordingly, a need exists for a method for optimal placement of data associated with visualization of massive multi-attribute data in pixel bar charts. A
5 need also exists for a method for presenting an overview of the data while also presenting the detailed information for each data item. A need also exists for a method for presenting data in a manner for detecting close relationships and trends among the data. A need also exists for a pixel visualization technique that presents data in a categorized manner that is easy to understand.

For Release

SUMMARY OF THE INVENTION

A method for placement of data for visualization of multidimensional data sets using multiple pixel bar charts is presented. Data is received comprising a plurality of records, each record having a plurality of attributes. From the plurality of attributes, a set of attributes is determined for placement of the plurality of records in a graphically displayable array comprising a plurality of data points, each data point representing one record of the plurality of records. In one embodiment, the graphically displayable array is a pixel bar chart. The plurality of records are arranged to construct the pixel bar chart for presenting the data in a format for detecting relationships between the plurality of records. The present invention provides a pixel visualization technique having a method of optimal data placement.

In one embodiment, the plurality of records are sorted by a first dividing attribute corresponding to horizontal axis, and partitioned into groups according to the first dividing attribute. In one embodiment, the plurality of records are then sorted by a second dividing attribute corresponding to vertical axis, and partitioned into sub-groups according to the second dividing attribute. The records of each of each sub-group are then sorted according to a first ordering attribute and a second ordering attribute. A visual indicator (e.g., a color) is then applied to each record according to a visual indicator attribute.

The present invention provides a method of efficient pixel placement for producing dense pixel visualizations that are capable of showing large volumes of multi-dimensional data sets and provides a heuristic placement method to resolve locality and ordering constraints. The present invention allows for the partition of the data set according to the x-axis and y-axis and for placement of pixels of each partition in the corresponding regions according to the x-and y- ordering. As such, the present invention provides a method for generating a pixel bar chart for detecting close relationships and trends among the data that is categorized and easy to understand.

HP-10014772-1/JPH/MJB

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention:

5

FIGURE 1 is a flowchart diagram illustrating steps in a process for generating a pixel bar chart in accordance with an embodiment of the present invention.

FIGURE 2 is a flowchart diagram illustrating steps in a process for placement
10 of data for visualization of multidimensional data sets using multiple pixel bar charts in accordance with an embodiment of the present invention.

FIGURE 3a is a block diagram illustrating the partitioning of records into
15 groups accordingly in accordance with an embodiment of the present invention.

15

FIGURE 3b is a block diagram illustrating the sorting of a plurality of records within a group by a second dividing attribute and partitioning the records into sub-groups accordingly in accordance with an embodiment of the present invention.

20

FIGURE 3c is a block diagram illustrating the arranging of a plurality of records within a sub-group by a first ordering attribute and a second ordering attribute in accordance with an embodiment of the present invention.

FIGURE 4 is a flowchart diagram illustrating steps in a process for pixel placement within a sub-group of a pixel bar chart in accordance with an embodiment of the present invention.

5

FIGURE 5 is an illustration of an exemplary sub-group 500 in accordance with one embodiment of the present invention.

For Filing

DETAILED DESCRIPTION

In the following detailed description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be apparent to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are not described in detail in order to avoid obscuring aspects of the present invention.

Some portions of the detailed descriptions which follow are presented in terms of procedures, steps, logic blocks, processing, and other symbolic representations of operations on data bits within a computer memory. These descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. A procedure, computer executed step, logic block, process, etc., is here and generally conceived to be a self-consistent sequence of steps of instructions leading to a desired result. The steps are those requiring physical manipulations of data representing physical quantities to achieve tangible and useful results. It has proven convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers or the like.

20

Portions of the present invention are comprised of computer-readable and computer executable instructions which reside, for example, in computer-usable media

of a computer system. It is appreciated that the present invention can operate within a number of different computer systems including general purpose computer systems, embedded computer systems, and stand alone computer systems specially adapted for controlling automatic test equipment.

5

Figure 1 is a flowchart diagram illustrating steps in a process 100 for generating a pixel bar chart in accordance with an embodiment of the present invention. Steps of process 100, in the present embodiment, may be implemented with any computer languages used by those of ordinary skill in the art. In one
10 embodiment, process 100 is for graphically presenting and visually mining large volumes of data using a graphically displayable array.

In one embodiment, the graphically displayable array is a pixel bar chart. In one embodiment, each record is represented by one unique pixel of a display. In
15 another embodiment, each record is represented by a data point of a display, wherein the data point comprises a plurality of pixels.

At step 110 of process 100, data comprising a plurality of records is received. Each record comprises a plurality of attributes, wherein each attribute corresponds to
20 a particular item of information of the record. For example, a consumer electronics business has a record for each order the business handles. In this example, each record may have attributes corresponding to the order number, the price of the order, the

customer identification number (ID), and the quantity of items ordered, as well as other attributes corresponding to customer data. It should be appreciated that each attribute of each record represents a value of the represented data item. As such, references to an attribute (e.g., first dividing attribute) with respect to a record are
5 intended to refer to the corresponding value for that attribute of the record.

At step 120, a set of attributes is determined from the plurality of attributes, wherein the set of attributes is for placement of the plurality of records in the graphically displayable array. In one embodiment, a first dividing attribute, a second
10 dividing attribute, a first ordering attribute, a second ordering attribute and a visual indicator attribute are selected.

In one embodiment, the first dividing attribute corresponds to a horizontal axis and the second dividing attribute corresponds to a vertical axis of the graphically
15 displayable array. The first dividing attribute and second dividing attribute are for partitioning the data into groups and sub-groups. It should be appreciated that the first dividing attribute and the second dividing attribute are distinct attributes selected from the plurality of attributes. However, each of the first dividing attribute and the
20 second dividing attribute may be the same as the first ordering attribute, the second ordering attribute and the visual indicator attribute.

In one embodiment, the first ordering attribute corresponds to a horizontal axis and the second ordering attribute corresponds to a vertical axis of the graphically displayable array. The first ordering attribute and second ordering attribute are for placing the individual records within each of the groups and sub-groups. It should be appreciated that the first ordering attribute and the second ordering attribute are distinct attributes selected from the plurality of attributes. However, each of the first dividing attribute and the second dividing attribute may be the same as the first dividing attribute, the second dividing attribute and the visual indicator attribute.

In one embodiment, the visual indicator attribute corresponds to a visual indicator. In one embodiment, the visual indicator is a color for pixel coloring. It should be appreciated that the visual indicator attribute is selected from the entire plurality of attributes. As such, the visual indicator attribute may be the same as any of the other selected attributes.

At step 130 of process 100, the plurality of records are arranged to construct the graphically displayable array for presenting said data in a format for detecting relationships between the plurality of records. In one embodiment, the present invention provides a process for placement of data for visualization of multidimensional data sets using multiple pixel bar charts in accordance with an embodiment of the present, as described in process 200 of Figure 2.

The present invention provides a method of efficient pixel placement for producing dense pixel visualizations that are capable of showing large volumes of data sets and provides a heuristic placement method to resolve locality and ordering constraints. The present invention allows for the partition of the data set according to the x-axis and y-axis and for placement of pixels of each partition in the corresponding regions according to the x-and y- ordering. As such, the present invention provides a method for generating a pixel bar chart for detecting close relationships and trends among the data that is categorized and easy to understand.

Figure 2 is a flowchart diagram illustrating steps in a process 200 for placement of data for visualization of multidimensional data sets using multiple pixel bar charts in accordance with an embodiment of the present invention. Steps of process 200, in the present embodiment, may be implemented with any computer languages used by those of ordinary skill in the art. In one embodiment, process 200 is for graphically presenting and visually mining large volumes of data using a graphically displayable array.

At step 210 of process 200, the plurality of records are sorted by the first dividing attribute. In one embodiment, the first dividing attribute is selected for partitioning the plurality of records (e.g., the data) into groups. In one embodiment, the groups correspond to a partitioning of the screen of a display according to the horizontal axis (e.g., x-axis).

At step 220, the records are partitioned into groups according to the first dividing attribute. In one embodiment, the first dividing attribute corresponds to the vertical axis (D_x).

5

Figure 3a is a block diagram of a graphically displayable array 300 (e.g., a pixel bar chart) as shown on a display illustrating the partitioning of records into groups 306 accordingly in accordance with an embodiment of the present invention.

10

Graphically displayable array 300 has horizontal axis (e.g., x-axis) 302 and vertical axis (e.g., y-axis) 304 and comprises a plurality of data points (e.g., pixels). The records are sorted by the first dividing attribute, which corresponds to horizontal axis 302. The records are divided into groups 306 according to each record's value of the first dividing attribute.

15

At step 230, the records are partitioned into sub-groups according to the second dividing attribute. In one embodiment, the second dividing attribute corresponds to the vertical axis (D_y). It should be appreciated that in an embodiment of the present invention, no attribute is specified for the second dividing attribute.

20

Figure 3b is a block diagram of a graphically displayable array 320 (e.g., a pixel bar chart) as shown on a display illustrating the partitioning of records within a group

306 into sub-groups 308 accordingly in accordance with an embodiment of the present invention.

Graphically displayable array 320 has horizontal axis (e.g., x-axis) 302 and
5 vertical axis (e.g., y-axis) 304 and comprises a plurality of data points (e.g., pixels).
The records are divided into groups 306 according to each record's value of the first
dividing attribute. Within each group 306, the records are divided into sub-groups 308
according to each record's value of the second dividing attribute.

10 At step 240, the records of each sub-groups (or group if no second dividing
attribute is selected) are arranged and placed according to the first ordering attribute
and the second ordering attribute. In one embodiment, the arrangement and placement
of pixels is conducted according to a process for pixel placement within a sub-group
of a pixel bar chart as described in process 400 of Figure 4. In one embodiment, the
15 first ordering attribute corresponds to the horizontal axis (O_x) and the second ordering
attribute corresponds to the vertical axis (O_y).

Figure 3c is a block diagram of graphically displayable array 340 (e.g., a pixel
bar chart) as shown on a display illustrating the arranging of a plurality of records
20 within a sub-group 308 by a first ordering attribute and a second ordering attribute in
accordance with an embodiment of the present invention.

Graphically displayable array 340 has horizontal axis (e.g., x-axis) 302 and vertical axis (e.g., y-axis) 304 and comprises a plurality of data points (e.g., pixels). In one embodiment, the records are divided into groups 306 according to each record's value of the first dividing attribute (e.g., no second dividing attribute is determined).

- 5 The records of each group 306 are placed according to a first ordering attribute and a second ordering attribute. Ordering arrows 310 and 312 indicate the direction of record placement within each group 306.

- 10 In another embodiment, the records are divided into sub-groups 308 according to each record's value of the first dividing attribute and value of the second dividing attribute. The records of each sub-group 308 are placed according to a first ordering attribute and a second ordering attribute. Ordering arrows 310 and 312 indicate the direction of record placement within each sub-group 308.

- 15 At step 250, a visual indicator is applied to each record according to the visual indicator attribute. In one embodiment, the visual indicator is a color. In one embodiment, non-linear 256 RGB (red-green-blue) color scale is used for determining the color for each record. In another embodiment, a non-linear gray-scale color scale is used to determine the color for each record. It should be appreciated that any color
- 20 scale or range, both linear and non-linear, may be used in regard to the present invention. In one embodiment, the value of the visual indicator attribute is normalized

to the range 0 to 1. In one preferred embodiment, the normalization is nonlinear.

Then the range 0 to 1 is mapped to a color range.

In one embodiment, where multiple pixel bar charts are used, different
 5 attributes are the visual indicator attribute in different bar charts, which enables a user
 to relate the different coloring attributes and detect partial relationships among the
 records. In the present embodiment, the dividing and ordering attributes have to stay
 in the same order:

$$10 \quad \langle D_x, D_y, O_x, O_y, C \rangle$$

where $D_x, D_y, O_x, O_y, C \in \{A_1, \dots, A_k\} \cup \perp$ and D_x/D_y are the dividing attributes in x-
 /y-direction, O_x/O_y are the ordering attributes in x-/y-direction, and C is the coloring
 attribute. The element \perp is used if no attribute is specified.

15

Figure 4 is a flowchart diagram illustrating steps in a process 400 for pixel
 placement within a sub-group of a pixel bar chart in accordance with an embodiment
 of the present invention. The pixel placement within one sub-group requires a two-
 dimensional sort. The present invention provides a heuristic process for pixel
 20 placement by partitioning the data set into subsets according to the first ordering
 attribute (O_x) and the second ordering attribute (O_y).

At step 410, the one-dimensional histograms for O_x and O_y are determined for providing an efficient pixel placement within a single sub-group. The one-dimensional histograms for O_x and O_y are used to determine the ∂ -quantiles of O_x and O_y . In one embodiment, if the sub-group under consideration has extension $w \times h$ pixels, the

5 $1/w, \dots, (w-1)/w$ -quantities for the partitioning of O_x and the $1/h, \dots, (h-1)/h$ -quantities for the partitioning of O_y are determined. The quantiles are then used to determine the partitions X_1, \dots, X_w of O_x and Y_1, \dots, Y_h of O_y . The partitions X_1, \dots, X_w are sorted according to O_y and the partitions Y_1, \dots, Y_h according to O_x .

10 At step 420, the pixel in the lower-left corner is placed (e.g., position (1,1), of the sub-group). In one embodiment, the pixel is placed according to Equation 1:

Equation 1

$$f^{-1}(1,1) = \left\{ d_s \mid \min_{d_s \in X_1} \{d_s, a_2\} = \min_{d_s \in Y_1} \{d_s, a_1\} \right\}$$

15

At step 430, all pixels in the lower pixel row and the left pixel row of the sub-group are placed. In one embodiment, the lower pixel row is placed according to Equation 2:

20

Equation 2

$$f^{-1}(i, 1) = \left\{ d_s \mid \min_{d_s \in X_i} \{d_s, a_2\} \right\} \quad \forall i = 1..w$$

In one embodiment, the left pixel row is placed according to Equation 3:

5

Equation 3

$$f^{-1}(1, j) = \left\{ d_s \mid \min_{d_s \in Y_j} \{d_s, a_1\} \right\} \quad \forall j = 1..h$$

At step 440, the remaining pixels are iteratively placed in the sub-group. In one embodiment, the pixels are placed starting from the lower left of the sub-group to the and ending at the upper right of the sub-group. In one embodiment, if pixels at positions $(i-1, j)$ and $(i, j-1)$ are already placed, the pixel at position (i, j) is determined according to Equation 4:

15

Equation 4

$$f^{-1}(i, j) = \left\{ d_s \mid \min_{d_s \in X_i \cap Y_j} \{d_s, a_1 + d_s, a_2\} \right\}$$

if $X_i \cap Y_j \neq \emptyset$

In one embodiment, where the data structures have been placed as described in step 410, the pixel to be placed at position can be determined in $O(1)$ time if

$X_i \cap Y_j \neq \emptyset$. If $X_i \cap Y_j = \emptyset$, the partitions X_i and Y_j must be iteratively extended and $d_s \in (X_i \cup X_{i+1}) \cap Y_j$ be considered.

In one embodiment, if this set is still empty, $d_s \in (X_i \cup X_{i+1}) \cap (Y_j \cup Y_{j+1})$ must
5 be considered, and so on, until a data point to be placed is found. It should be appreciated that process 400 is efficient due to the data structure used.

Figure 5 is an illustration of an exemplary sub-group 500 in accordance with one embodiment of the present invention. It should be appreciated that sub-group
10 500 may be a group in instances where no second dividing attribute is selected. Sub-group 500 comprises a plurality of pixels. In one embodiment, the plurality of pixels are pixels of a display screen (e.g., a computer monitor).

Lower-left pixel 502 (e.g., position (1,1)) is placed first according to step 420
15 of process 400 (Figure 4). Lower pixel row 504 (e.g., position (i,1)) and left pixel row 506 (e.g., position (1,i)) are placed next, according to step 430 of process 400. Remaining pixels 508 (e.g., position (i+1, i+1),) are then placed in sub-group 500, according to step 440 of process 400.

20 The preferred embodiment of the present invention a method for placement of data for visualization of multidimensional data sets using multiple pixel bar charts, is

thus described. While the present invention has been described in particular embodiments, it should be appreciated that the present invention should not be construed as limited by such embodiments, but rather construed according to the below claims.

HP-10014772-1/JPH/MJB